Consequence-Based Approach-Avoidance Training:

A New and Improved Method for Changing Behavior

Pieter Van Dessel, Sean Hughes, and Jan De Houwer

Ghent University, Belgium

Word count: 1.995

Author Note

PVD, SH, JDH, Department of Experimental Clinical and Health Psychology, Ghent University. PVD is supported by a Postdoctoral fellowship of the Scientific Research Foundation, Flanders (FWO-Vlaanderen). JDH is supported by Methusalem Grant BOF16/MET_V/002 of Ghent University. Correspondence concerning this article should be sent to Pieter.vanDessel@UGent.be

Abstract

The repeated performance of approach or avoidance actions in response to specific stimuli (e.g., alcoholic drinks) is often considered a most promising type of cognitive bias modification that can reduce unwanted behavior (e.g., alcohol consumption). Unfortunately, approach-avoidance training sometimes fails to produce desired outcomes (e.g., in the context of unhealthy eating). We introduce a novel training task in which approach-avoidance actions are followed by affective consequences. Four experiments (total N = 1547) found stronger changes in voluntary approach-avoidance behavior, implicit and explicit evaluations and consumer choices for consequence-based approach-avoidance training in the food domain. Moreover, this novel type of training reduced self-reported unhealthy eating behavior after a 24-hour delay and unhealthy snacking in a taste test. Our results contrast with dominant (association-formation) accounts of approach-avoidance training effects and support an inferential explanation. They further suggest that consequence-based approach-avoidance training, and inference training more generally, holds promise for the treatment of clinical behavior.

Keywords: approach-avoidance training, consequences, unhealthy food consumption, cognitive bias modification, inferential theory

Consequence-Based Approach-Avoidance Training: A New and Improved Method for Changing Behavior

Approach and avoidance represent two fundamental classes of behaviors that organisms have at their disposal when interacting with the environment. When faced with a positive (or appetitive) stimulus, it is often beneficial to approach that stimulus whereas it is usually beneficial to avoid negative (or aversive) stimuli. Many theories assume that, as a result of this evolutionary benefit, evaluative processing is closely tied to approach-avoidance behavior (Lang, Bradley, & Cuthbert, 1990; Solarz, 1960; Strack & Deutsch, 2004). More specifically, these theories postulate that affective evaluation of a stimulus automatically predisposes people to approach or avoid that stimulus (Chen & Bargh, 1999; but see Rotteveel & Phaf, 2004). It is further assumed that these approach-avoidance tendencies can represent a cognitive bias that mediates unwanted or maladaptive behavior. For instance, a strong tendency to approach appetitive but unhealthy stimuli (e.g., unhealthy foods) can facilitate unhealthy behavior (e.g., consuming such foods) (Hofmann, Gschwendner, Friese, Wiers, & Schmitt, 2008).

A growing number of researchers have tried to modify maladaptive behaviors such as unhealthy eating by establishing changes in approach-avoidance tendencies via approach-avoidance (AA) training. In a typical AA training task, participants repeatedly perform approach or avoidance actions in response to specific stimuli. For instance, in studies on addiction, participants consistently avoid appetitive stimuli they might normally approach (e.g., alcoholic drinks) by moving away from it or by moving it away from them, and approach control stimuli (e.g., water). Several studies indicate that AA training can be effective in the treatment of clinical conditions. For instance, avoidance training might have beneficial effects in the treatment of addiction (e.g., alcohol-dependence: Wiers, Eberl, Rinck, Becker, & Lindernmeyer, 2011, or

smoking addiction: Wittekind, Feist, Schneider, Moritz, & Fritzsche, 2015), and approach training might be effective for treating social anxiety (Taylor & Amir, 2012), spider phobia (Jones, Vilensky, Vasey, & Fazio, 2013), or depression (Becker et al., 2016). Yet, several studies have also failed to find such effects. For instance, training people to avoid unhealthy foods is often ineffective for changing unhealthy eating behavior (e.g., Becker, Jostmann, Wiers, & Holland, 2015). Two meta-analyses even concluded that there is little reliable evidence for the effectiveness of AA training interventions on (many) clinical outcomes (Cristea, Kok, & Cuijpers, 2015, 2016, but see Kakoschke, Kemps, & Tiggemann, 2017).

A recent review further established that, overall, data do not support dominant explanations of AA training effects which assume that behavioral AA tendencies reflect mental stimulus-response associations that are gradually changed on the basis of repeated stimulus-action pairings (Jones, Hardman, Lawrence, & Field, 2017; see also Spruyt et al., 2013). In light of this conclusion, we recently developed an alternative model which postulates that inferential processes underlie AA training effects (Van Dessel, Hughes, & De Houwer, 2018). From this perspective, repeated performance of AA actions in response to a stimulus (e.g., avoidance of alcoholic drinks) leads to the formation of inferences about (evaluative) properties of the stimulus (e.g., that alcohol is to-be-avoided). This inferential learning can then affect subsequent stimulus-related actions (e.g., alcohol consumption).

The inferential account assumes that AA training effects depend on specific boundary conditions. One such condition is whether participants infer information about the consequences that result from responding in a particular way. It is well-established that learning about positive or negative action consequences determines the performance of related actions (Thorndike, 1905), and that stimulus-based actions such as approach-avoidance are typically facilitated when

more positive action consequences are anticipated (Eder & Hommel, 2013). It is possible that affective consequences of approach-avoidance actions are sometimes learned during AA training. For instance, participants who approach feared stimuli (e.g., spiders: Jones et al., 2013) might learn that the approach responses do not lead to the anticipated negative consequences. Critically, however, typical AA training does not specify in a clear manner what action consequences participants should learn, which might explain null effects and even contrast effects in past work. For instance, repeated avoidance of desired stimuli, like chocolate, might sometimes be unpleasant (e.g., because chocolate look tasty) which could hinder rather than facilitate future avoidance of the stimuli (Becker et al., 2015, Experiment 3).

With this in mind, we performed four experiments that tested the effectiveness of a novel type of AA training in which approaching or avoiding specific stimuli consistently led to positive or negative consequences. In Experiment 1, one group of participants performed typical AA training in which they consistently approached products of one unknown food brand and avoided products of another by moving either a manikin or avatar representing themselves towards or away from the products. Another group performed consequence-based AA training in which they approached and avoided products from both brands with an avatar. Importantly, for one brand, approach always produced positive consequences and avoidance always produced negative consequences, whereas the actions produced opposite consequences for the other brand. Action consequences were chosen such that they would facilitate evaluative learning in the food context (i.e., improve or decline the avatar's general health). In Experiment 2, we further 'super-charged' consequence-based AA training by making action consequences relevant to task goals (i.e., participants were instructed to try and maximize avatar health). On the basis of our inferential account, we predicted that this would facilitate the inferential step from action performance to

evaluative learning and therefore enhance AA training effects. Both experiments probed for effects on consumer choices, voluntary approach-avoidance responses, implicit (i.e., automatic) and explicit (i.e., controlled) stimulus evaluations.

Experiments 3 and 4 extended our investigation to familiar healthy and unhealthy foods and examined the impact of consequence-based AA training on (a) self-reported healthy eating behavior and intentions (completed 24 hours after the intervention; Experiment 3) and (b) the amount of unhealthy food participants consumed in an ad libitum snack task (Experiment 4).

Method

Participants and Design

A total of 600, 525, and 420 volunteers participated online in Experiments 1, 2, and 3 via the Prolific Academic website (https://prolific.ac). Participants in Experiment 4 were 184 undergraduate students from Ghent University. The sample size of all experiments was determined on the basis of an *a priori* power analysis such that the sample size would provide sufficient power (i.e., power > 0.80) to detect a small to medium effect. Prior to data-collection, target sample size was pre-registered together with the study design, data-analytic plans, and experimental hypotheses on the Open Science Framework website. For all experiments and all measures, we predicted that consequence-based AA training would lead to stronger effects than typical AA training or control training. The pre-registered plans, raw data, experimental and analytic scripts are available at https://osf.io/3anqx/. After data-exclusion on the basis of pre-registered criteria (see SOM-R for details), we retained the data of 519 (300 women, mean age = 34, *SD* = 12), 455 (288 women, mean age = 34, *SD* = 12), 389 (219 women, mean age = 34, *SD* = 13), and 184 participants (59 women, mean age = 20, *SD* = 2). A total of 307 participants (78.9%) completed the second part of Experiment 3 (mean delay = 29 hours, *SD* = 4).

Procedure of Experiments 1 and 2

AA training task. After providing informed consent and completing demographic information, participants performed one of three different versions of the AA training task.

Manikin task (Experiments 1-2). The manikin task was adopted from Woud, Maas, Becker, and Rinck (2013). It was selected as a typical AA training task for this study because it has produced robust effects in the past (see SOM-R). In this task, participants performed 80 trials in which they saw a stick figure (manikin) that represented themselves along with a product from one of two novel food brands (named Vekte and Empeya) (Figure 1). Depending on the color of the frame, participants approached the product by moving the manikin towards it or avoided the product by moving the manikin away from it. Products from one brand (e.g., Vekte) were always surrounded by the colored frame that had to be approached and products from the other brand (e.g., Empeya) were always surrounded by the colored frame that had to be avoided. Whenever the participant made a correct response by pressing the up or down key on the keyboard, the manikin moved towards (up) or away (down) from the food product.

Avatar task (Experiment 1). The avatar task was designed for the purpose of this study and served as a typical AA training control for the consequence-based AA training task described below. Before starting the task, participants selected whether a male or female avatar would represent them in the task. Training consisted of 80 trials in which participants first saw the avatar standing in front of a fridge. The fridge then gradually opened until a brand product with colored frame appeared. A correct response (pressing the up or down key) resulted in an avatar movement towards or away from the brand product.

Avatar consequences task (Experiments 1-2). On each trial of this task, participants saw the avatar, the fridge, and the product with the frame, but also a health bar that was presented

above the avatar. After a correct response and resulting avatar movement, action consequences were presented such that (1) the health bar gradually depleted, the sentence ‘I feel sick’ appeared, and the avatar had an unhealthier appearance (negative consequences) or (2) the health bar filled, the sentence ‘I feel healthy’ appeared, and the avatar had a healthier appearance (positive consequences). Products from both brands were presented equally often with blue and green frames (and, thus, were approached and avoided an equal number of times). Crucially, however, approaching one brand always produced positive consequences and avoiding it produced negative consequences, whereas approaching the other brand always produced negative consequences and avoiding it produced positive consequences.

Goal-relevant avatar consequences task (Experiment 2). This task was designed to “supercharge” the avatar consequences task by making action consequences relevant for participants’ task goals. Participants were told that each time they would approach or avoid the products they would see the avatar become more healthy or sick and that their task would be to make the avatar as healthy as possible by performing these actions. During the task, there were no colored frames surrounding the brand products and participants freely selected whether to move the avatar towards or away from the brand product. Contingencies between products, actions, and action consequences were the same as in the avatar consequences task.

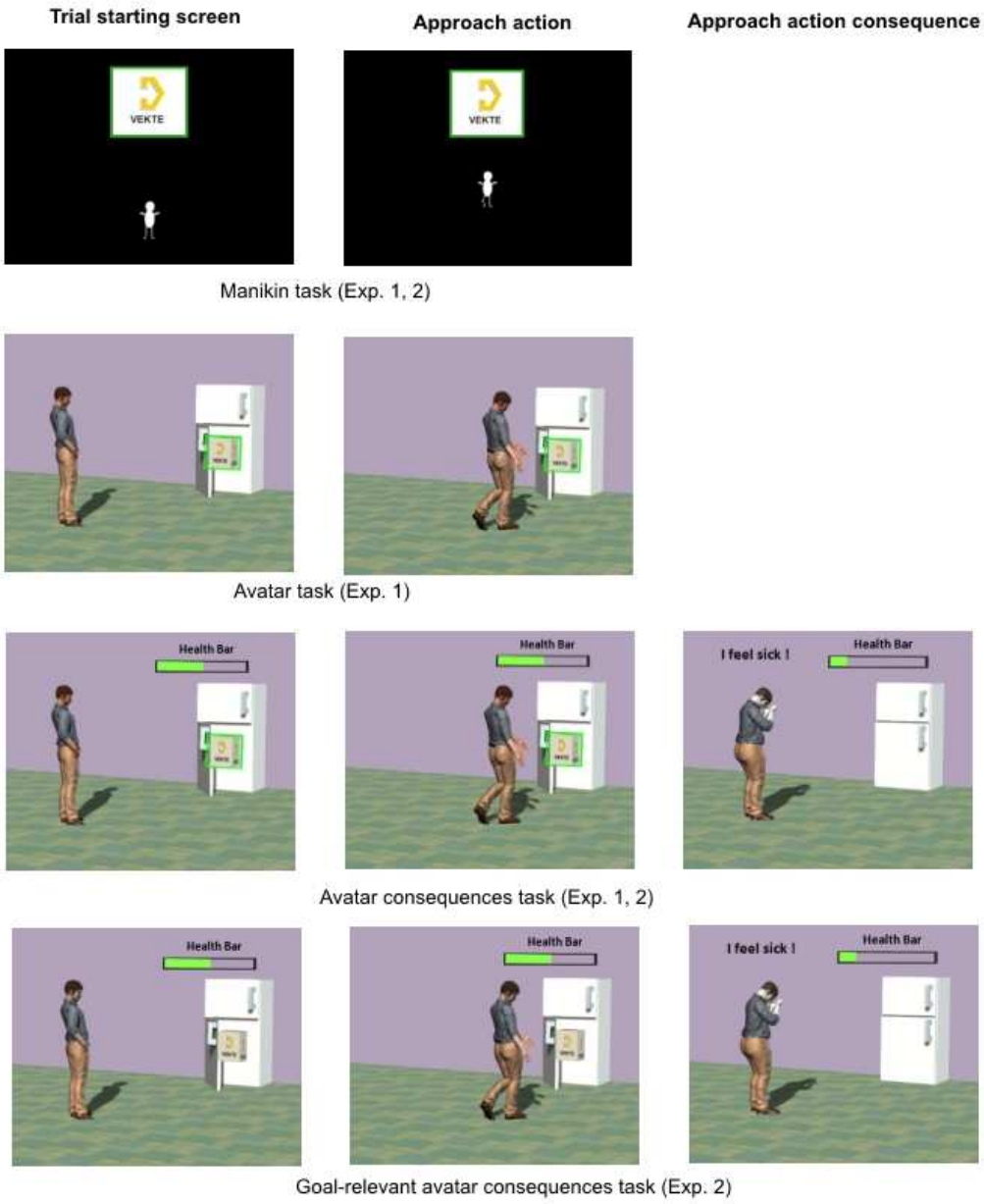

Figure 1. Illustration of a trial with an approach response to a Vekte brand product in the four task conditions of Experiments 1 and 2.

Outcome measures. After the AA training, participants completed a question that probed consumer behavior. They then completed an Implicit Association Task (IAT; Greenwald, McGhee, & Schwartz, 1998) that measured implicit evaluations of the two food brands and an

explicit rating task that measured explicit evaluations of those same brands. Finally, they completed a task that measured their voluntary approach-avoidance behavior.

Consumer choices. Participants were informed that we would be willing to send them a free sample of products from the two food brands and they were then asked to indicate whether they would prefer products from Vekte, Empeya, both, or neither.

Implicit evaluations (IAT). The IAT of Experiment 1 was constructed following the recommendations of Nosek, Greenwald, and Banaji (2005). Participants categorized eight attribute words (e.g., wonderful, evil) as 'positive' or 'negative' and four different versions of the brand logos as their respective names ('Vekte' or 'Empeya'). In two experimental blocks of 56 trials each, stimuli related to one brand and positive shared a response key and stimuli related to the other brand and negative shared a second response key. The IAT of Experiment 2 was personalized, with the category labels 'I like' and 'I dislike' used to categorize the attribute words (see Han, Olson, & Fazio, 2006).

Explicit evaluations. Participants indicated how positive or negative they considered each of the two brands by using a Likert scale ranging from 1 (*very negative*) to 9 (*very positive*).

Approach-avoidance behavior. Participants were told that they would perform a final task in which they would again see the brand products and the manikin (or avatar). They were asked to imagine that they were now at home and that they were free to choose which action to make when they encountered the products (i.e., approach or avoid). Participants then completed 10 trials of the same AA training task they had completed before. However, there were no colored frames surrounding the food products and participants were free to either approach or avoid the products without any consequences of doing so.

Exploratory questions. We also probed the extent to which participants (1) had learned the correct contingencies between food brands and approach-avoidance actions (and action consequences in the consequence task conditions), (2) had imagined that they were the manikin (or avatar), (3) liked the action of approaching or avoiding in general, and (4) had provided evaluative responses in order to comply with experimental demands (demand compliance), or to react against these demands (reactance).

Procedure of Experiments 3 and 4

Phase 1. Procedures were similar to Experiment 2 with the following exceptions. First, we used known healthy and unhealthy food products (e.g., carrots and cookies) as stimuli rather than products from novel food brands. Second, when participants started the experiment they were asked to indicate the extent to which they (1) had the goal to eat healthy (healthy eating intention), (2) felt hungry at that moment (hunger), (3) often ate healthy (healthy eating behavior), and (4) found it difficult to cut down or stop eating unhealthy foods (healthy eating behavior difficulty). Third, participants either performed a manikin task in which they approached and avoided healthy and unhealthy foods an equal number of times (*control condition*) or a goal-relevant avatar consequences task in which approaching healthy foods and avoiding unhealthy foods always led to positive health outcomes while avoiding healthy foods and approaching unhealthy foods always led to negative health outcomes (*goal-relevant avatar consequences task condition*) (Figure 2). Experiment 3 additionally included a typical AA training task in which participants always approached healthy foods and avoided unhealthy foods with a manikin (*manikin task condition*).

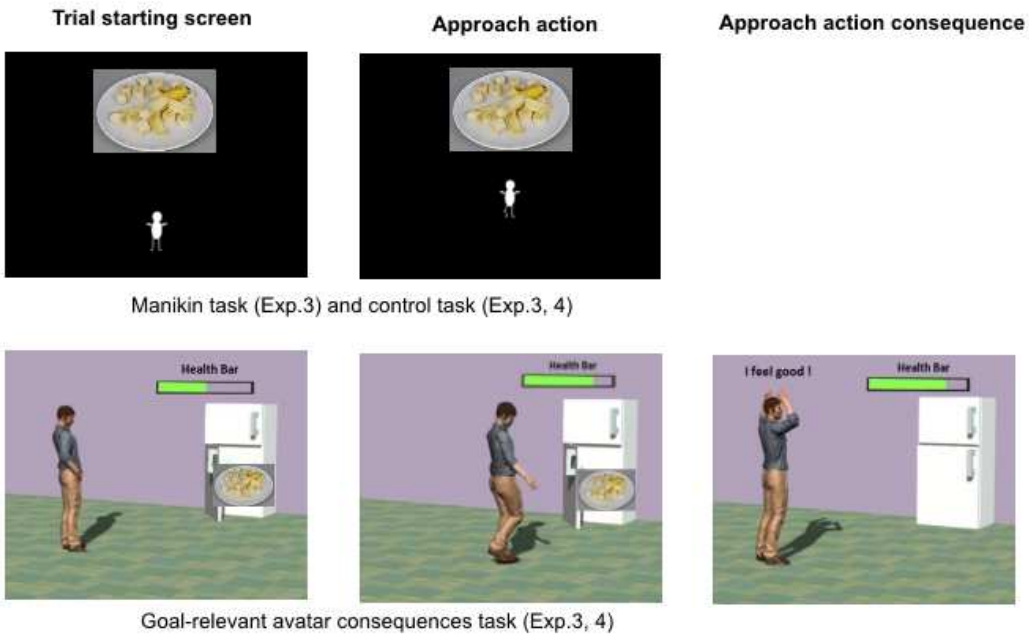

Figure 2. Illustration of a trial with an approach response to a healthy food product in the three task conditions of Experiments 3 and 4.

Fourth, the outcome measures in Experiment 3 included a consumer choice task, explicit evaluation rating task, pIAT, and an AA task that consisted of 12 free-choice AA trials. Participants in Experiment 4 did not perform the free-choice AA task but they completed an ad-libitum snack task, adapted from Haynes, Kemps, and Moffit (2015), in which participants were presented with four full bowls of pre-weighed popular energy-dense snack foods: mixed candy (3.3 kcal/g), potato chips (5.3 kcal/g), M&Ms (4.7 kcal/g), and cheese flips (5.1 kcal/g). Participants were instructed to rate the foods on different sensory characteristics for use in an unrelated study. After providing these ratings, participants were informed that they could consume as much of these snacks as they wanted. Unbeknownst to participants, bowls were weighed at the end of the experiment. Fifth, the consumer choice task consisted of 6 trials in which participants indicated which of two food products they would prefer to receive a coupon

for (Experiment 3) or actually receive after experiment completion (Experiment 4). Trials either presented trade-off pairs involving an unhealthy food and a less attractive healthy food (e.g., chocolate cookie and rice cake) or controlled pairs involving an unhealthy food and an equally attractive healthy food (e.g., banana and waffle). Finally, at the end of Experiment 4, participants reported task engagement for the AA training phase (short version of the Dundee Stress State Questionnaire; Helton & Näswall, 2015) and temptation to eat the foods presented in the snack task (7-point Likert scale: 1 = not at all; 7 = extremely).

Phase 2 (Experiment 3). Participants were contacted via the Prolific Academic website one day after completing the first experimental phase. Upon return, they completed four questions that probed healthy and unhealthy eating behavior, healthy eating behavior intention, and difficulty experiences to stop eating unhealthily, in the period between the first and second study phase. A final question asked participants to what extent they currently had the goal to eat healthily (healthy eating goal). All answers were provided on Likert scales ranging from 1 to 9.

Results

Results of Experiments 1 and 2

Consumer choices. Behavioral choice scores were computed by recoding responses to the consumer behavior question such that -1 indicates that participants selected only the *negative brand* (i.e., the brand that was consistently avoided in typical AA training or for which approach produced negative and avoidance produced positive consequences in consequence-based AA training), 0 indicates that participants selected both or neither brands, and 1 indicates that participants selected only the *positive brand* (i.e., the brand that was consistently approached in typical AA training or for which approach produced positive and avoidance produced negative consequences in consequence-based AA training). Scores were subjected to a Kruskal-Wallis

test, which revealed significant main effects of AA Task Condition for both experiments, χ^2s > 14.65, *ps* < .001. Participants selected the positive brand more often than the negative brand in all conditions of Experiments 1 (Table 1) and 2 (Table 2), *ps* < .003, d_{zs} > 0.41. Crucially, participants selected the positive brand more often in the avatar consequences condition than in the manikin or avatar conditions, *ps* < .003. Moreover, the positive brand was selected more often in the goal-relevant avatar consequences condition of Experiment 2 than in both other conditions, *ps* < .035. See SOM-R for more detailed results of this and all other analyses.

Implicit and explicit evaluations. IAT scores indicated an implicit preference for the positive over the negative brand in all task conditions of both experiments, *p*s < .001, d_zs > 0.42. ANOVA's revealed a main effect of AA Task Condition in Experiment 2, $F(2,443) = 9.02$, $p < .001$, but not in Experiment 1, $F(2,507) = 2.07$, $p = .13$. In Experiment 2, IAT scores were higher in the goal-relevant avatar consequences condition than in manikin or avatar consequences conditions, *t*s > 2.75, *p*s < .007. IAT scores did not differ significantly between the latter two conditions, $t(246) = 1.12$, $p = .27$.

Explicit ratings indicated a preference for the positive over the negative brand in all task conditions of both experiments, *p*s <.001, d_zs > 0.74. ANOVA's revealed main effects of AA Task Condition in both experiments, *F*s > 11.60, *p*s < .001. In Experiment 1, scores were higher in the avatar consequences condition than in manikin or avatar conditions, *t*s > 3.79, *p*s <.001. Experiment 2 also found higher scores in the avatar consequences condition than in the manikin condition, $t(246) = 4.85$, $p < .001$. Furthermore, scores were higher in the goal-relevant avatar consequences condition than in either of the other two conditions, *t*s > 5.60, *p*s < .001.

Approach-avoidance behavior. Participants more often approached the positive than the negative brand in all task conditions of both experiments, *p*s <.001, d_zs > 0.54. ANOVA's on

approach-avoidance behavior scores revealed main effects of AA Task Condition for both experiments, Fs > 3.02, ps < .050. In both experiments, scores were higher in the avatar consequences condition than in manikin or avatar conditions, ts > 1.81, ps < .070. In Experiment 2, scores were also higher in the goal-relevant avatar consequences condition than in both other conditions, ts > 5.14, ps < .001.

Results of Experiments 3 and 4

Consumer choices. Participants in all task conditions selected healthy foods more often than unhealthy foods, ps < .001, d_zs > 0.35, except for the control condition in Experiment 4 (Table 3). Behavioral choice scores were subjected to an ANOVA with AA Task Condition and Choice (Trade-off, Control) as between-subject factors. For this and all subsequent analyses we included pre-rated Health Behavior, Health Intention, Health Behavior Difficulty, and Hunger as covariates if they significantly improved model fit. We observed main effects of Choice, Health Behavior, Health Intention, and Hunger, Fs > 5.30, ps < .023, and, crucially, also a main effect of AA Task Condition, Fs > 8.91, ps < .004. In both experiments, behavioral choice scores were higher for the goal-relevant avatar consequences condition compared to the control and manikin conditions, ts > 2.70, ps < .008. Scores did not differ significantly between manikin and control condition (Experiment 3), $t(261) = 1.84$, $p = .066$.

Implicit and explicit evaluations. Participants in all task conditions exhibited an implicit preference for healthy foods over unhealthy foods, ps < .001, d_zs > 1.95. The ANOVA on IAT scores revealed main effects of Health Intention and IAT Block Order, Fs > 9.40, ps < .003, and, crucially, also the main effect of AA Task Condition, Fs > 3.83, ps < .028. In both experiments, IAT scores were higher in the goal-relevant avatar consequences condition than in the control

condition and manikin conditions, *t*s > 2.16, *p*s < .032. Scores did not differ significantly between manikin and control condition (Experiment 3), $t(261) = 0.41$, $p = .68$.

Explicit ratings indicated an explicit preference for healthy foods over unhealthy foods in all task conditions, *p*s <.001, d_zs > 0.98. The ANOVA on explicit rating scores revealed main effects of Health Intention, Health Behavior, and Hunger, *F*s > 5.84, *p*s < .016. We also observed a main effect of AA Task Condition in Experiment 3, $F(2,382) = 3.08$, $p = .047$, but not in Experiment 4, $F(1,180) = 0.06$, $p = .81$. In Experiment 3, scores were higher in the goal-relevant avatar consequences condition compared to the control condition, $t(254) = 2.45$, $p = .015$, but not compared to the manikin condition, $t(257) = 1.57$, $p = .12$, and scores did not differ significantly between manikin and control condition, $t(261) = 0.89$, $p = .38$.

Approach-avoidance behavior (Experiment 3). Participants in all task conditions approached more healthy than unhealthy foods, *p*s <.001, d_zs > 0.67. The ANOVA on approach-avoidance behavior scores revealed main effects of Health Behavior and Health Intention, *F*s > 6.19, *p*s < .014, as well as a main effect of AA Task Condition, $F(2,382) = 5.25$, $p = .006$. Compared to the control condition, scores were higher for both the goal-relevant avatar consequences condition and the manikin condition, *t*s > 2.24, *p*s < .026. Scores did not differ significantly between goal-relevant consequences and manikin condition, $t(257) = 0.91$, $p = .36$.

Snack eating (Experiment 4). Mean snack intake (in grams) was subjected to an ANOVA with AA Task Condition as between-subjects factor and pre-rated Health Intention and Hunger as covariates. We observed main effects of Health Intention, and Hunger, *F*s > 4.29, *p*s < .040, and, crucially, also a main effect of AA Task Condition, $F(1,180) = 5.67$, $p = .018$. Participants' snack intake was lower in the goal-relevant avatar consequences condition compared to the control condition.

Phase 2 questions (Experiment 3). A multivariate ANOVA on Phase 2 ratings revealed effects of Health Behavior, Health Behavior Difficulty, and Health Intention, *F*s > 11.07, *p*s < .001, and most importantly also a main effect of AA Task Condition, $F(10,301) = 4.46$, $p < .001$ (Table 4). Follow-up analyses that included Health Behavior, Health Behavior Difficulty, and Health Intention as covariates revealed a significant effect of AA Task Condition for unhealthy eating behavior ratings and healthy eating behavior intention ratings, *F*s > 3.31, *ps* < .038, and a marginally significant effect for healthy eating goal ratings, $F(2,301) = 2.76$, $p = .065$. Participants in the goal-relevant avatar consequences condition provided lower ratings for unhealthy eating behavior compared to participants in the control and manikin conditions, *t*s > 2.23, *p*s < .027, and higher ratings for healthy eating behavior intention and healthy eating goal compared to participants in the control condition, *t*s > 2.29, *p*s < .023. Ratings did not differ significantly between manikin and control condition, *t*s < 0.99, *p*s > .33.

Discussion

Four experiments examined the effects of a novel type of approach-avoidance (AA) training in which approach and avoidance responses to food products were consistently followed by positive or negative consequences. In Experiment 1, consequence-based AA training had a bigger impact on consumer choices, approach-avoidance behavior, and explicit (but not implicit) evaluation of novel food brands than typical AA training. Experiment 2 replicated these findings and showed that goal-relevant action consequences enhanced the effects of consequence-based AA training (also on implicit evaluation). Experiment 3 found that consequence-based AA training also produced bigger effects than typical AA training in the context of healthy and unhealthy foods. Moreover, compared to a control training, consequence-based AA training (but not typical AA training) reduced self-reported unhealthy eating behaviors and increased healthy

eating intentions 24 hours after training. Experiment 4 further showed that consequence-based AA training reduces actual unhealthy eating as measured in a snack task.

The fact that consequence-based AA training produced robust effects on consumer choices, approach-avoidance behavior, and implicit and explicit evaluations, is important given the inconsistent effects that typical AA training often produces (Cristea et al., 2015; 2016), especially in the context of food products (Becker et al., 2015). Although we also observed clear effects of typical AA training (mainly in the context of novel foods), effects of consequence-based AA training were consistently stronger. Importantly, consequence-based AA training also seemed to reduce actual unwanted behavior such as the consumption of unhealthy foods (i.e., participants reported less unhealthy eating behavior in the consequence-based AA training condition of Experiment 3 and consumed less unhealthy foods in the snack task of Experiment 4).

The current findings fit with an inferential account of AA training effects which asserts that these effects result from inferences related to goal-directed action (Van Dessel et al., 2018). In consequence-based AA training, participants directly learn about the consequences of AA responses to certain (food) stimuli and this may have caused them to anticipate similar outcomes for similar actions (e.g., actual unhealthy eating). Responses for which positive outcomes are anticipated generally have a higher value and are therefore facilitated in comparison to responses for which negative outcomes are anticipated (Eder, Rothermund, De Houwer, & Hommel, 2015). Because inferential learning is assumed to be determined by activated goals, AA training effects are further enhanced when action consequences are goal-relevant. Typical AA training also allows participants to learn about the consequences of AA responses. For instance, avoiding unhealthy foods might facilitate retrieval of information consistent with the idea that unhealthy foods are to-be-avoided (e.g., negative consequences of eating unhealthy). Yet, unlike

consequence-based AA training, typical AA training does not specify nor require that action consequences are learned, which might explain why effects of typical AA training were smaller than effects of consequence-based AA training.

Our results contrast with dominant accounts of AA training effects which assume that experienced contingencies between stimuli and approach-avoidance responses cause an automatic re-wiring of cognitive biases based on mental stimulus-response associations that mediates these effects (e.g., Wiers et al., 2011, 2013). Because typical AA training involves stronger stimulus-response contingencies than consequence-based AA training, these accounts predict stronger changes in cognitive biases and resulting effects for typical compared to consequence-based AA training. Of course, the current results do not preclude the possibility that non-inferential (e.g., associative) mechanisms contribute to AA training effects. Some have argued that (only) automatic (e.g., unintentional) AA training effects depend on associative mechanisms (e.g., Kawakami et al., 2007). However, automaticity should not be conflated with underlying processes (e.g., inferential reasoning can produce automatic effects). In fact, our inferential account assumes that AA training involves an important automatization component to the extent that the repeated nature of the task facilitates more automatic inferences. Moreover, there was no indication that our consequence-based AA training effects are more likely to be based on controlled processes than typical AA training effects. For instance, though it is possible that demand characteristics biased AA training effects (see Sharpe & Whelton, 2016), we found that typical AA training effects correlated more strongly with demand compliance ratings than consequence-based AA training effects (see SOM-R). Moreover, Experiment 4 showed AA training effects (e.g., on IAT scores) in the absence of effects on more controlled, explicit liking ratings.

Note that caution is still warranted when interpreting the current findings. First, it is possible that other factors than learned consequences might explain observed dissociations between typical and consequence-based AA training effects. For instance, general task attention or task engagement might be enhanced in consequence-based AA training tasks which could strengthen effects. However, this explanation does not fit with the observation that (1) participants' overall AA training task performance was better in the typical AA training task conditions and (2) self-reported task engagement did not mediate AA training effects (see SOM-R). Second, effects on actual (unwanted) behavior were only established in the snack eating task of Experiment 4 (compared to a control condition). However, the fact that four high-powered experiments showed strong effects of consequence-based AA training on many outcomes (including self-reported unhealthy eating after a one-day delay) does lead us to believe that consequence-based AA training has practical use.

We therefore hope that future research will examine whether beneficial results of consequence-based AA training can also be obtained in clinical samples and in other (clinical) domains (e.g., alcohol consumption, depression). Such studies could also examine moderators of consequence-based AA training effects. We already found evidence that goal-relevance of action consequences might be one important moderator, as predicted by our inferential account (Van Dessel, Hughes, & De Houwer, 2018). However, this account also predicts that other task adaptations may further improve effects (e.g., including a higher number of training trials/sessions, using more lifelike or personally relevant affective consequences, or training across multiple contexts). Our results also open the door for designing interventions that aim to facilitate adaptive inferences ("inference training") on the basis of other actions than approach and avoidance. This approach bears resemblance to current "nudging" interventions that aim to

modify behavior by providing (subtle) environmental cues (Benartzi et al., 2017) and to effective therapies used in clinical practice that target beliefs underlying maladaptive behavior (i.e., cognitive behavioral therapy: Beck & Dozios, 2011). However, the fact that participants need to derive new information themselves and repeatedly act upon it is different from current treatments and might facilitate (automatic) effects (see Wiers et al., 2011, for evidence in the context of AA training). Inference training (via AA training) might also be easier to distribute (e.g., in mobile apps) and to incorporate into existing initiatives (e.g., health promotion interventions).

Author contributions

All authors were involved in developing the study concept and contributed to the design. Data collection and data-analyses were performed by P. Van Dessel. P. Van Dessel drafted the manuscript. J. De Houwer and S. Hughes provided critical revisions. All authors approved the final version of the manuscript for submission.

References

Beck, A. T., Dozios, D. J. A. (2011). Cognitive therapy: Current status and future directions. *Annual Review of Medicine, 62,* 397–409. doi:10.1146/annurev-med-052209-100032.

Becker, E.S., Ferentzi, H., Ferrari, G., Möbius, M., Brugman, S. Custers, J., Geurtzen, N., Wouters, J., & Rinck, M. (2016). Always Approach the Bright Side of Life: A General Positivity Training Reduces Stress Reactions in Vulnerable Individuals. *Cognitive Research and Therapy, 40,* 57-71.

Becker, D., Jostmann, N. B., Wiers, R. W., & Holland, R. W. (2015). Approach avoidance training in the eating domain: Testing the effectiveness across three single session studies. *Appetite, 85,* 58–65. doi: 10.1016/j.appet.2014.11.017.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25,* 215–224. doi:10.1177/0146167299025002007

Cristea, I. A., Kok, R. N., & Cuijpers, P. (2015). Efficacy of cognitive bias modification interventions in anxiety and depression: Meta-analysis. *British Journal of Psychiatry, 206,* 7-16.

Cristea, I. A., Kok, R. N., & Cuijpers, P. (2016). The effectiveness of cognitive bias modification interventions for substance addictions: A meta-analysis. *PLoS ONE,* 11.

Eder, A.B., & Hommel, B. (2013). Anticipatory control of approach and avoidance: An ideomotor approach. *Emotion Review, 5,* 275-279

Eder, A.B., Rothermund, K., De Houwer, J., & Hommel, B. (2015). Directive and incentive functions of affective action consequences: an ideomotor approach. *Psychological Research, 79*, 630-649.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–80. doi: 10.1037/0022-3514.74.6.1464

Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally created extrapersonal associations on the Implicit Association Test. Journal of Experimental Social Psychology, 42, 259 –272.

Haynes, A., Kemps, E., & Moffitt, R. (2015). Inhibitory self-control moderates the effect of changed implicit food evaluations on snack food consumption. *Appetite, 90*, 114–122.

Helton, W. S., & Näswall, K. (2015). Short Stress State Questionnaire: Factor structure and state change assessment. *European Journal of Psychological Assessment, 31*(1), 20-30.doi:10.1027/1015-5759/a000200

Hofmann, W., Gschwendner, T., Friese, M., Wiers, R.W., & Schmitt, M. (2008). Working memory capacity and self-regulatory behavior: Toward an individual differences perspective on behavior determination by automatic vs. controlled processes. *Journal of Personality and Social Psychology, 95*, 962–977.doi:10.1037/a0012705

Jones, A., Hardman, C. A., Lawrence, N., & Field, M. (2017). Cognitive training as a potential treatment for overweight and obesity: A critical review of the evidence. *Appetite.* doi:10.1016/j.appet.2017.05.032

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: evidence regarding insects and spiders. *Emotion, 13*5, 989–996. doi: 10.1016/S0896-6273(02)00858-9

Kakoschke, N., Kemps, E., & Tiggemann, M. (2017). Approach bias modification training and consumption: A review of the literature. *Addictive Behaviors, 64,* 21-28.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*, 377–395. doi:10.1037/0033-295X.97.3.377.

Nosek, B. A., Greenwald, A. G., Banaji, M. R. (2005) Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166–180.

Rotteveel, M. & Phaf, R.H. (2004). Automatic affective evaluation does not automatically predispose for arm flexion and extension. *Emotion, 4*, 156-172.

Sharpe, D., & Whelton, W. J. (2016). Frightened by an Old Scarecrow: The remarkable resilience of demand characteristics. *Review of General Psychology*.

Solarz A. K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of Experimental Psychology, 59,* 239–245. doi:10.1037/h0047274.

Spruyt A., De Houwer J., Tibboel H., Verschuere B., Crombez G., Verbanck P., et al. (2013). On the predictive validity of automatically activated approach/avoidance tendencies in abstaining alcohol-dependent patients. *Drug and Alcohol Dependence, 127*, 81–8610.1016/j.drugalcdep.2012.06.019

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220-247. doi: 10.1207/s15327957pspr0803_1

Taylor, C. T., & Amir, N. (2012). Modifying automatic approach action tendencies in individuals with elevated social anxiety symptoms. *Behaviour Research and Therapy, 50,* 529–536. doi: 10.1016/j.brat.2012.05.004

Van Dessel, P., Eder, A. B., & Hughes, S. (in press). Mechanisms Underlying Approach-Avoidance Training Effects on Stimulus Evaluation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

Van Dessel, P., Hughes, S., & De Houwer, J. (2018). How Do Actions Influence Attitudes? An Inferential Account of the Impact of Action Performance on Stimulus Evaluation. Manuscript invited for revision at Personality and Social Psychology Review. Prepint available at: https://osf.io/kb3wq/

Wiers, R. W., Eberl, C., Rinck, M., Becker, E., Lindenmeyer, J. (2011). Re-training automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science, 22,* 490-497. doi:10.1177/0956797611400615

Wiers, R. W., Gladwin, T. E., Hofmann, W., Salemink, E., & Ridderinkhof, K. R. (2013). Cognitive bias modification and control training in addiction and related psychopathology: Mechanisms, clinical perspectives and ways forward. *Clinical Psychological Science, 1*, 192-212.

Wittekind, C. E., Feist, A., Schneider, B. C., Moritz, S., & Fritzsche, A. (2015). The approach-avoidance task as an online intervention in cigarette smoking: A pilot study. *Journal of*

Behavior Therapy and Experimental Psychiatry, *46*, 115–120. doi: 10.1016/j.jbtep.2014.08.006

Woud, M. L., Maas, J., Becker, E. S., & Rinck, M. (2013). Make the manikin move: Symbolic approach–avoidance responses affect implicit and explicit face evaluations. *Journal of Cognitive Psychology*, *25*, 738–744. doi: 10.1080/20445911.2013.817413

Tables

Table 1. *Means and effect sizes of all outcome measures for participants in the three task conditions of Experiment 1.*

	Manikin task N = 173			Avatar task N = 171			Avatar consequences task N =175		
	Mean (SD)	95% CI	d_z	Mean (SD)	95% CI	d_z	Mean (SD)	95% CI	d_z
Consumer choice	0.31 (0.54)	[0.22,0.39]	0.56	0.33 (0.60)	[0.24,0.42]	0.54	0.50 (0.68)	[0.40,0.60]	0.73
Explicit rating	1.76 (2.90)	[1.28,2.25]	0.61	1.99 (3.03)	[1.50,2.48]	0.66	3.31 (3.75)	[2.83,3.80]	0.88
IAT score	0.28 (0.45)	[0.22,0.35]	0.62	0.28 (0.47)	[0.20,0.33]	0.59	0.20 (0.48)	[0.13,0.27]	0.42
Approach-avoidance	1.55 (2.80)	[1.13,1.97]	0.55	1.71 (2.90)	[1.27,2.16]	0.59	2.29 (3.21)	[1.85,2.73]	0.71

Table 2. *Means and effect sizes of all outcome measures for participants in the three task conditions of Experiment 2.*

	Manikin N = 157			Avatar consequences N = 146			Goal-relevant avatar consequences N =152		
	Mean (SD)	95% CI	d_z	Mean (SD)	95% CI	d_z	Mean (SD)	95% CI	d_z
Consumer choice	0.24 (0.58)	[0.15,0.33]	0.42	0.47 (0.62)	[0.36,0.57]	0.75	0.63 (0.51)	[0.54,0.71]	1.22
Explicit rating	1.75 (2.38)	[1.29,2.22]	0.74	3.49 (3.68)	[3.01,3.97]	0.95	5.60 (2.71)	[5.13,6.07]	2.07
IAT score	0.28 (0.38)	[0.22,0.33]	0.72	0.23 (0.44)	[0.16,0.29]	0.52	0.40 (0.37)	[0.34,0.46]	1.07
Approach-avoidance	1.53 (2.61)	[1.13,1.93]	0.59	2.58 (3.09)	[2.16,2.99]	0.83	4.09 (1.77)	[3.69,4.50]	2.32

Table 3. *Means and effect sizes of all Phase 1 outcome measures for participants in the three task conditions of Experiments 3 and 4.*

		Manikin N = 133 (Exp. 3)			Control N = 130 (Exp. 3) / 92 (Exp. 4)			Goal-relevant avatar consequences N = 126 (Exp. 3) / 92 (Exp. 4)		
		Mean (SD)	95% CI	d_z	Mean (SD)	95% CI	d_z	Mean (SD)	95% CI	d_z
Consumer choice	Experiment 3	0.47 (0.83)	[0.33,0.60]	0.57	0.28 (0.80)	[0.15,0.42]	0.36	0.75 (0.84)	[0.60,0.90]	0.90
	Experiment 4				0.09 (0.68)	[-0.05,0.23]	0.14	0.38 (0.80)	[0.21,0.54]	0.47
Explicit rating	Experiment 3	2.82 (2.53)	[2.42,3.21]	1.08	2.57 (2.63)	[2.17,2.96]	0.98	3.27 (2.71)	[2.87,3.68]	1.23
	Experiment 4				4.15 (2.09)	[3.78,4.53]	1.95	4.22 (2.15)	[3.84,4.59]	1.99
IAT score	Experiment 3	0.80 (0.34)	[0.74,0.85]	2.38	0.78 (0.37)	[0.72,0.84]	2.11	0.90 (0.35)	[0.85,0.96]	2.58
	Experiment 4				0.83 (0.37)	[0.77,0.89]	2.12	0.92 (0.25)	[0.86,0.99]	3.50
Approach-avoidance	Experiment 3	2.56 (2.70)	[2.13,2.99]	0.92	1.86 (2.82)	[1.43,2.29]	0.67	2.84 (2.52)	[2.41,3.28]	1.16
Snack eating	Experiment 4				46.7 (33.5)	[40.8,52.1]	/	36.5 (25.2)	[31.2,42.4]	/

Table 4. *Mean score ratings of the Phase 2 questions for participants in the three task conditions of Experiment 3.*

	Manikin N = 108		Control N = 99		Goal-relevant avatar consequences N =100	
	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI
Healthy eating behavior	5.24 (1.93)	[4.93,5.54]	5.28 (1.88)	[4.97,5.60]	5.53 (2.02)	[5.21,5.84]
Unhealthy eating behavior	4.68 (1.89)	[4.37,4.98]	4.63 (1.78)	[4.32,4.94]	4.13 (1.79)	[3.82,4.45]
Healthy eating intention	6.02 (2.29)	[5.68,6.35]	5.78 (2.28)	[5.43,6.12]	6.41 (2.07)	[6.06,6.76]
Difficulty to stop eating unhealthily	4.88 (2.27)	[4.52,5.23]	4.80 (2.21)	[4.43,5.16]	4.98 (2.28)	[4.62,5.35]
Healthy eating goal	6.67 (1.97)	[6.42,6.92]	6.54 (1.96)	[6.28,6.80]	6.96 (1.56)	[6.71,7.22]